American Community Survey Design and Methodology (January 2014)

Chapter 15: Improving Data Quality by Reducing Non-Sampling Error

United States™ Census Bureau

Version 2.0
January 30, 2014

[This page intentionally left blank]

Table of Contents

[This page intentionally left blank]

Chapter 15: Improving Data Quality by Reducing Non-Sampling Error

15.1 Overview

Total survey error includes two components: sampling error and nonsampling error. Chapters 4 and 11 provide information about the steps the Census Bureau takes to reduce American Community Survey (ACS) sampling error. As with all surveys, the quality of ACS data reflects how well the data collection and processing procedures address potential sources of nonsampling error, including coverage, nonresponse, and measurement errors, as well as errors that may arise during data capture and processing. Groves (1989) identified four primary sources of nonsampling error:

- Coverage Error: The failure to give some units in the target population any chance of selection into the sample, or giving units more than one chance of selection.
- Nonresponse Error: The failure to collect complete data from all units in the sample.
- Measurement Error: The inaccuracy in responses recorded on survey instruments, arising from:
  - The effects of interviewers on the respondents' answers to survey questions.
  - Respondents' inability to answer questions, lack of requisite effort to obtain the correct answer, or other psychological or cognitive factors.
  - Faulty wording of survey questions.
  - Data collection mode effects.
- Processing Error: Errors introduced after the data are collected, including:
  - Data capture errors.
  - Errors arising during coding and classification of data.
  - Errors arising during editing and item imputation of data.

This chapter identifies the operations and procedures designed to reduce these sources of non-sampling error and thus improve the quality of ACS estimates. It also includes information about ACS Quality Measures that provide an indication of the potential for some types of nonsampling error. Finally, it describes the annual review process used to demonstrate that ACS data meet the Census Bureau's Statistical Quality Standards. The ACS releases the survey estimates, as well as the Quality Measures, at the same time each year so that users can consider data quality in conjunction with the survey estimates. The Quality Measures for years 2000 to currently released data are located on the ACS Quality Measures Website: <http://www.census.gov/acs/www/methodology/sample_size_and_data_quality/>.

Additional data products describing data quality are available through the American FactFinder (AFF): <http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml>.

15.2 Coverage Error

All surveys experience some degree of coverage error. It can take the form of undercoverage or overcoverage. Undercoverage occurs when units in the target population do not have a chance of selection into the sample, for example, when our address frame is missing certain housing units or when respondents erroneously exclude some people from a household roster. Overcoverage occurs when units or people have multiple chances of selection, for example, when our address frame lists a housing unit more than once. In general, coverage error will affect survey estimates if the characteristics of the individuals or units excluded or included in error differ from the characteristics of those correctly listed.

The ACS Quality Measures contain housing-level and person-level coverage rates as indicators of the potential for coverage error. The ACS calculates coverage rates for the total resident population by sex at the national, state, and Puerto Rico geographies, and at the national level only for Hispanics and non-Hispanics crossed by the five major race categories: White, Black, American Indian and Alaska Native, Asian, and Native Hawaiian and Other Pacific Islander. The total resident population includes persons in both housing units (HUs) and group quarters (GQ). In addition, these measures include a coverage rate specific to the GQ population at the national level. We calculate coverage rates for HUs at the national and state level, with the exception of Puerto Rico because independent HU estimates are not available.

The coverage rate is the ratio of the ACS population or housing estimate of an area or group to the independent estimate for that area or group, multiplied by 100. The Census Bureau uses independent data on housing, births, deaths, immigration, and other categories to produce official estimates of the population and HUs each year. The base for these independent estimates is the decennial census counts. We weight the numerator in the coverage rates to reflect the probability of selection into the sample, subsampling for personal visit follow-up, and adjustments for unit nonresponse. The weight used for this purpose does not include poststratification adjustments (weighting adjustments that make the weighted totals match the independent estimates), since the control totals used in production for this purpose serve as the basis for comparison for the coverage rates.

Over-and undercoverage can be partially adjusted as part of the poststratification process, that is, adjusting weights to independent population control totals. The ACS corrects for potential overcoverage or undercoverage by controlling to these official estimates on specific demographic characteristics and at specific levels of geography. As the coverage rate for a particular subgroup drops below 100 percent (indicating undercoverage), the weights of its members are adjusted upward in the final weighting procedure to reach the independent estimate. If the rate is greater than 100 percent (indicating overcoverage), the weights of its members are adjusted downward to match the independent estimates. Chapter 11 provides more details regarding the ACS weighting process.

The ACS uses the Master Address File (MAF) as its sampling frame, and includes several procedures for reducing coverage error in the MAF, described below. Chapter 3 provides further details.

- Twice a year, the Census Bureau receives the U.S. Postal Service (USPS) Delivery Sequence File (DSF) that consists of the addresses including a house number and street name rather than a rural route or post-office box. Geography Division uses this file to update the city-style addresses on the MAF.
- The ACS nonresponse follow-up operation provides ongoing address and geography updates.
- The Community Address Updating System (CAUS) can provide address updates (as a counterpart to the DSF updates) that cover predominately rural areas where city-style addresses generally are not used for mail delivery. The Census Bureau chose to put the CAUS program on hold in late 2006 due to the address canvassing operation for the 2010 Census and restarted the program in 2010. The CAUS program in 2013 plans to list approximately 2,000 blocks.

The ACS conducts a telephone follow-up operation on mail responses with conflicting information about the total number of household members and on all households with six or more persons. This follow up confirms that the ACS includes the appropriate persons, reducing the possible impact of coverage error. The Internet instrument includes an error message to a respondent when they provide conflicting information on total persons, and allows up to 20 household members.

15.3 Nonresponse Error

There are two forms of nonresponse error: unit nonresponse and item nonresponse. Unit nonresponse results from the failure to obtain the minimum required data from a unit in sample. Item nonresponse occurs when respondents do not report individual data items, or provide data considered invalid or inconsistent with other answers.

Surveys strive to increase both unit and item response to reduce the potential for bias introduced into survey estimates. Bias results from systematic differences between the nonrespondents and the respondents. Without data on the nonrespondents, surveys cannot easily measure differences between the two groups.

Unit Nonresponse

The Census Bureau presents survey response and nonresponse rates as part of the ACS Quality Measures. The survey response rate is the ratio of the units interviewed after data collection to the estimate of all units that were eligible to be interviewed. The ACS Quality Measures provide separate rates for HUs and GQ persons. For the HU response rate, the numerator is a weighted estimate of the number of cases that were interviewed after all modes of data collection.

To accurately measure unit response, the ACS estimates the universe of cases eligible to be interviewed, which becomes the denominator of the unit response rate.

The ACS Quality Measures also include the weighted estimates of nonresponse rates broken down by the reason for nonresponse. These reasons include refusal, unable to locate the sample unit, no one home during the data collection period, temporarily absent during the interview period, language problem, insufficient data (not enough data collected to consider it a response), and other (such as ‘‘sample address not accessible’’; ‘‘death in the family’’; or cases not followed up due to budget constraints).

For the GQ person response rate, the numerator is a weighted estimate of all interviewed GQ persons after personal visit. The denominator is the weighted estimate of the total number of persons eligible to be interviewed in GQs. For the GQ rates, there are two additional reasons for 4 noninterviews: whole GQ refusal and whole GQ other (such as unable to locate the GQ).

Item Nonresponse

The ACS Quality Measures provide information about item nonresponse. When respondents do not report individual data items, or provide data considered invalid or inconsistent with other answers, the Census Bureau imputes the necessary data. The imputation methods use either rules to determine acceptable answers (referred to as ‘‘assignment’’) or answers from similar people or HUs (‘‘allocation’’). Assignment involves logical imputation, in which a response to one question implies the value for a missing response to another question. For example, first name often can be used to assign a value to sex. Allocation involves using statistical procedures to impute for missing values.

The ACS Quality Measures include summary allocation rates as a measure of the extent to which item nonresponse required imputation. Beginning with the 2007 ACS data (including multiyear data), two allocation rates—overall HU characteristic and overall person characteristic allocation rates—are available on the AFF at the nation and state level (plus the county level for five-year periods). Allocation rates of many individual characteristics are available on the ACS Quality Measures Web site at the national and state level for 2000 to the present. In addition, the ACS releases imputation tables on AFF that allow users to compute allocation rates for published variables and all published geographies.

The ACS reduces the potential for nonresponse bias by reducing the amount of unit and item nonresponse through procedures and processes listed below.

- Response to the ACS is mandated by law, and information about the mandatory requirement to respond is provided in most materials and reinforced in communication with respondents in all stages of data collection.
- The ACS survey operations include two stages of nonresponse follow-up: a computer-assisted telephone interview (CATI) follow-up for Internet and mail nonrespondents, and

a computer-assisted personal interview (CAPI) follow-up for a sample of remaining nonrespondents and unmailable addresses cases.

- The ACS mailing protocol is based on a strategy shown in research studies to obtain a high response rate (Dillman, 1978; Tancreto et al., 2012, Matthews et al., 2012): a prenotice letter, an invitation to respond online with a message on the envelope stating that response is ‘‘required by law,’’ a postcard reminder, a paper questionnaire mailing for nonrespondents to the initial mailing, and, for those same nonrespondents, a second reminder postcard.
- If we cannot find a telephone number for an address that did not respond by Internet or mail at the point when operations switch to CATI nonresponse follow-up, a third postcard is sent urging response by mail, Internet or by calling the toll-free ACS Telephone Questionnaire Assistance (TQA) help line.
- The mailing package includes a frequently asked questions (FAQ) motivational brochure explaining the survey, its importance, and its mandatory nature.
- Both the online survey and paper questionnaire use designs that reflect accepted principles of respondent friendliness and navigation. The intent of the design is to make it easier for respondents to navigate the survey, as well as provide cues for a valid response at an item level (such as providing examples of the type of response desired, or using a prefilled ‘0’ to indicate reporting monetary amounts rounded to the nearest dollar). Similarly, the Internet, CATI and CAPI instruments direct respondents and interviewers to ask the appropriate questions based on the respondent’s answers.
- The online survey and questionnaire provide a toll-free telephone number for respondents who have questions about the ACS in general or who need help in completing the survey.
- The ACS includes a telephone failed-edit follow-up (FEFU) interview with mail and Internet respondents whose responses indicate a discrepancy between the reported household size and the number of people for whom data are provided, and those who indicated a household size of six or more people. (The mail form allows data for only five people, so the FEFU operation collects data for any additional persons.) In addition, addresses from the Internet identified as a business or vacant are also sent to FEFU.
- The ACS uses permanent professional interviewers trained in refusal conversion methods for CATI and CAPI.
- Survey operations include providing support in other languages: the online survey is available in Spanish, a Spanish paper questionnaire is available on request, and there is a Spanish CATI/CAPI instrument. There are CATI and CAPI interviewers who speak Spanish and other languages as needed. Furthermore, we send a brochure in six languages (English, Spanish, Chinese, Korean, Russian, and Vietnamese) that provides a toll-free number for respondents to receive telephone questionnaire assistance in each language.

15.4 Measurement Error

Measurement error is defined as the difference between the recorded answer and the true answer. Measurement error may occur in any mode of data collection and can be caused by vague or ambiguous questions misinterpreted by respondents; questions that respondents cannot answer or questions where respondents deliberately falsify answers for social desirability reasons (see Tourangeau and Yan (2007) for information on social desirability); or interviewer characteristics or actions such as the tone used in reading questions, the paraphrasing of questions, or leading respondents to certain answers. In 2012, the Census Bureau conducted a reinterview with a subsample of ACS respondents to attempt to quantify the amount of inconsistency in reporting between the original interview and reinterview. This study, which will be available in early 2014, found minimal evidence of measurement error for most ACS questions.

The ACS minimizes measurement error in several ways, some of which also help to reduce nonresponse.

- ACS pretests new or modified survey questions in all modes before introducing them into the ACS as mandated in the Census Bureau Standard ‘‘Pretesting Questionnaires and Related Materials for Surveys and Censuses (Version 1.2),’’ available at: < http://www.census.gov/srd/pretest-standards.html >.
- The ACS uses a questionnaire design on the Internet and paper form that reflects accepted principles of respondent friendliness and navigation.[1]
- The ACS online survey provides help via a hyperlink for respondents who need additional information on how to interpret and respond to specific questions. Households that receive the paper ACS questionnaire also receive an instruction booklet that provides the same information.
- Respondents may call the toll-free TQA line and speak with trained interviewers for answers to general ACS questions or questions regarding specific items.
- The presentation of questions in the Internet, mail, CATI, and CAPI data collection modes reflect the strengths and limitations of each mode. For example, the online survey provides the ability to review/edit answers at the end of the survey, less complicated skip patterns on the mail questionnaire, breaking up questions with long or complicated response categories into separate questions for telephone administration, and including respondent flash cards for personal visit interviews.
- The Internet and CATI/CAPI instruments automatically navigate the questionnaire, showing only those questions appropriate for the interviewee based on their reported characteristics.

[1] http://www.census.gov/acs/www/Downloads/library/2012/2012_Tancreto_02.pdf

- The Internet and CATI/CAPI instruments include functionality that helps achieve valid responses for some questions. For example, the instruments check for values outside of the expected range to ensure that the reported answer reflects an appropriate response.
- Training for the permanent CATI and CAPI interviewing staff includes instruction on reading the questions as worded and answering respondent questions, and encompasses extensive role-playing opportunities. All interviewers receive a manual that explains each question in detail and provides detailed responses to questions often asked by respondents.
- Telephone interview supervisors and specially-trained staff monitor CATI interviews and provide feedback regarding verbatim reading of questions, recording of responses, interaction with respondents, and other issues.
- Field supervisors and specially-trained staff implement a quality reinterview program with CAPI respondents to minimize falsification of data.
- The Internet and CATI/CAPI instruments include a Spanish version, and bilingual CATI/CAPI interviewers provide language support in other languages.

Methods that make it easier for the respondent to understand the questions also increase the chances that the individual will respond to the questionnaire.

15.5 Processing Error

The final component of nonsampling error is processing error—error introduced in the post-data collection process of turning the responses into published data. For example, a processing error may occur in keying the data from the mail questionnaires. The miscoding of write-in responses, either clerically or by automated methods is another example. The degree to which imputed data differ from the truth also reflects processing error—specifically imputation error.

A number of practices are in place to control processing error (more details are discussed in Chapters 7 and 10). For example:

- Data capture of mail questionnaires includes a quality control procedure designed to ensure the accuracy of the final keyed data.
- Clerical coding includes a quality control procedure involving double-coding of a sample of the cases and adjudication by a third keyer.
- By design, automated coding systems rely on manual coding by clerical staff to address the most difficult or complicated responses.
- Procedures for selecting one interview or return from multiple returns for an address rely on a review of the quality of data derived from each response and the selection of the return with the most complete data.
- After completion of all four phases of data collection (Internet, mail, CATI, and CAPI), questionnaires with insufficient data do not continue in the survey processing, but instead receive a noninterview code and are accounted for in the weighting process.

- Edit and imputation rules reflect the combined efforts and knowledge of subject matter experts, as well as experts in processing, and include evaluation and subsequent improvements as the survey continues to progress.
- Subject matter and survey experts complete an extensive review of the data and tables, comparing results with previous years' data and other data sources.

15.6 Census Bureau Statistical Quality Standards

Beginning in 2012, the ACS program conducted a new, formal assessment to demonstrate that its data products meet the Census Bureau's recently released quality standards. Because the ACS covers a very broad set of topics, the Census Bureau chose a selection of seventeen important core 5-year estimates reflecting social, demographic, economic, and housing characteristics to gauge the quality of the survey's products. This assessment uses several of the ACS Quality Measures associated with nonsampling error including coverage rates, unit response rates, and item response rates (Sections 15.2 and 15.3 of this document provide an explanation about most of these types of nonsampling errors). We derived a combined nonsampling error rate from these three measures. Given the ACS goal to produce estimates for small areas, counties were chosen to assess quality for smaller geographies that are large enough to result in stable quality measures. The assessment calculated these four rates at the county level, summarizing those results as medians. To meet the quality standard: the median county-level unit response rate must equal or exceed 60 percent, the median county-level coverage rates must equal or exceed 70 percent and the median county-level item response rates for each one of the 17 estimates must equal or exceed 70 percent. The median county-level combined measure of nonsampling error (the product of a coverage rate, the unit response rate and the item response rate) must equal or exceed 50 percent for each of the 17 estimates.

The assessment also includes a measure of sampling error that does not come from the ACS Quality Measures. Sampling error in survey estimates arises due to the use of probability sampling. The ACS was designed to produce reliable tract-level estimates based on 5-year data. To determine a measure of sampling error, the assessment includes the median tract-level 5-year coefficient of variation (CV) for each of the 17 estimates. The CV is equal to the standard error (SE) of a weighted estimate divided by the weighted estimate itself. The majority of the estimates must have a CV less than or equal to 30 percent to meet the quality standard.

For both 2012 and 2013, the ACS data products have met the Census Bureau's Statistical Quality Standards. There is every indication that ACS will continue to meet them.

15.7 References

Biemer, P., and L. Lyberg. (2003) Introduction to Survey Quality, Hoboken, NJ: John Wiley and Sons.

ACS Design and Methodology Improving Data Quality by Reducing Nonsampling Error 15−5

U.S. Census Bureau

Dillman, D. (1978) Mail and Telephone Surveys: The Total Design Method, New York: John Wiley and Sons.

Groves, R. M. (1989) Survey Errors and Survey Costs, New York: John Wiley and Sons.

Groves, R. M., M. P. Couper, F. J. Fowler, J. M. Lepkowski, E. Singer, and R. Tourangeau. (2004) Survey Methodology, Hoboken, NJ: John Wiley and Sons.

Matthews, B., Davis, M, Tancreto, J.G., Zelenak, M.F, and Ruiter, M (2012) "2011 American Community Survey Internet Tests: Results from Second test in November 2011", U.S. Census Bureau Report.
http://www.census.gov/acs/www/Downloads/library/2012/2012_Matthews_01.pdf

Tourangeau, R., and T. Yan. (2007) ''Sensitive questions in surveys,'' Psychological Bulletin, 133(5): 859−883.

Tancreto, J.G., Zelenak, M.F., Davis, M, Ruiter, M., and Matthews, B. (2012) "2011 American Community Survey Internet Tests: Results from First Test in April 2011", U.S. Census Bureau Report. http://www.census.gov/acs/www/Downloads/library/2012/2012_Tancreto_01.pdf

Census Bureau (2002) ''Meeting 21st Century Demographic Data Needs—Implementing the American Community Survey: May 2002, Report 2: Demonstrating Survey Quality,'' Washington, DC.

Census Bureau (2004) ''Meeting 21st Century Demographic Data Needs—Implementing the American Community Survey: Report 7: Comparing Quality Measures: The American Community Survey's Three-Year Averages and Census 2000's Long Form Sample Estimates,'' Washington, DC.

Census Bureau (2006) ''Census Bureau Standard: Pretesting Questionnaires and Related Materials for Surveys and Censuses Version 1.2,'' Washington, DC.